Public Access to Digital Materials

Brewster Kahle, Internet Archive
Rick Prelinger, Prelinger Archives
Mary E. Jackson, Association of Research Libraries
Draft 0.4 March 5, 2001

The goal of universal access to our cultural heritage is within our grasp. With current digital technology we can build comprehensive collections, and with digital networks we can make these available to students and scholars all over the world. The current challenge is establishing the roles, rights, and responsibilities of our libraries and archives in providing public access to this information. With these roles defined, our institutions will help fulfill this epic opportunity of our digital age.

For the second time in history, people are laying plans to collect all information—the first time involved the Greeks which culminated in the Library of Alexandria, which is attributed with being the centerpiece of the world’s first university and illuminating human nature in a way never possible before. Now reports on how much total information there is (ref Lesk and Lyman) are leading many once again to take steps in building libraries that hold complete collections. The recent change that makes this possible is the move to digital technologies for storage, access, and coordinating distributed efforts.

While digital libraries are springing up in every corner, a recent experience may serve as an example of some of the advances and hurdles. Over the last five years the Internet Archive has built some of the largest text and moving images collections by developing cost-effective mechanisms of collection, cataloging, and preservation. By recording about 40 terabytes of text, 1000 movies, and television, much of the digital collecting technology has been successfully tested.

Some of the challenges encountered were issues of scale, funding, law, and access. The approach to the scale of the task has been to enlist partners and the newest computer technology to make this currently feasible-- and this technology gets better over time. Consequently, the technology has gotten to the point where scanning all books, digitizing all audio recordings, downloading all websites, and recording the output of all TV and radio stations is not only feasible but less costly than buying and storing the physical versions. As in traditional institutions, the funding for the Internet Archive has come through charitable giving, government funding, and in-kind donations. These funding sources form the support structure for most existing libraries so these challenges are understandable. Legal issues, however, remain the main area where more work is needed.

Some believe that the digital materials are fundamentally different from printed materials and have erected legal barriers to using the new materials in familiar ways. An article quoted the President of the American Library Association, Nancy Kranich, as saying that library-goers should be able to duplicate limited amounts of information for educational purposes. “Suppose you want to copy a journal article, quote a section of a book or use a line from a poem”, she says. "That is all permitted under the fair-use provision of the copyright law. In the digital arena, fair use has been narrowed to the point of disappearing… The publishing community does not believe that the public should have

To try to accommodate those interests, three proposals for a legal structure for public access to digital materials have emerged. The first proposal is the donation of intellectual property rights to preserves or conservancies in exchange for tax credits. The second proposal is to use interlibrary loan as a legal structure for the exchange of digital materials between libraries. The third embraces the lending library concept in which the public can get direct access to digital materials from their homes for a limited period.

The donation of the 1001 archival films by Prelinger Archives to the Internet Archive is a functional example of how a conservancy might provide widespread access to formerly unavailable cultural documents. While the donation approach is on fairly clear legal ground, it does not necessarily ensure comprehensive access to our cultural heritage because it relies on the good will of donors of digital content.

The existing interlibrary loan (ILL) system can offer a legal and organizational structure for rapidly expanding access to born-digital as well as digitized information. Interlibrary loan is an established mechanism to better serve the public through the pooling of collections to approximate a more complete library. While each exchange in the physical world costs over $25 and often a delay of two weeks or more, the digital equivalent does not need to be so expensive and slow.

The last approach, the loaning of digital materials, is specifically outlined in the Copyright Act of 1976 as a way to grant public access to television news recordings. A system to loan materials over the Internet will be demonstrated at the Internet Archive and Association of Research Libraries conference on Born Digital Materials in March 2001.

All of these approaches can be used in combination to great effect. Additionally, the cost of digitization and storage of our “analog” past can be shared among many libraries because of the linking of the computers via networks while at the same time maintaining control over the number of copies made. An Internet-based library system containing the majority of our audio, visual, printed, and digital cultural achievements would therefore cost less than the building of a new major library.

These approaches to building a library system of digital works can achieve the far-reaching goal of universal access to our cultural heritage. This paper will suggest how we can build this future.

**Vision for Public Access to Digital Materials**

Imagine a high school student in Singapore writing a report on the life of Madame Curie based on documentary film footage and original photographs from the Curie family's archives.

Imagine researching one’s family roots in Europe by browsing the original birth and marriage records from the "old country" from a PC in the library.
Imagine being diagnosed with cancer and having the world’s best medical research library and librarians less than one mile from your house.

Imagine a college student’s documentary film about her grandfather’s World War II battalion. This file could contain original military footage, current footage from towns involved in the battles, and interviews with surviving soldiers from the same battalion.

Anyone with access to a library can and should have access to this breadth of information. The digital divide between the information-age “haves” and “have-nots” can be overcome by making the majority of our culture’s information available to everyone. Our libraries already offer free public access to a portion of our cultural heritage, but any particular library has only a tiny fraction of our history, which often excludes most audio visual and digital materials. Worse yet, if libraries focus mostly on printed materials housed in physical libraries, then its essential role in society could diminish as society becomes more digital.

Digital technology can help bring the history of the world to every town and the history of every town to the world.

The current library and archive system has served the public well in the age of printed media. We now need to extend these models where possible and invent new models where needed to offer access to our digital cultural artifacts. Three sections follow that explain three methods of building public access.

**Intellectual Property Preserves and Conservancies**

To think about encouraging and sustaining public access to cultural resources is to consider basic questions of property and its privatization. Land use history offers precedents and a possible solution. In the late 19th and early 20th centuries, private corporations exerted unprecedented pressures on the "public domain" -- American land and natural resources. The aggressive pursuit of extractive interests such as mining, logging and agriculture threatened to exhaust public lands and encroach upon naturally or culturally significant sites. In response to this threat, the conservationist movement lobbied to organize a system of national forests, parks, and monuments. By preserving a limited public sphere not subject to the exercise of private property rights, the benefits of some wilderness and cultural sites were preserved for all.

In a similar way, preserves and conservancies are attempts to create and nurture public space within the complicated and intangible territory of intellectual property. Several people are presently proposing similar ideas. Eric Eldred is calling for a "copyright conservancy." [http://www.boston.com/globe/magazine/8-29/featurestory1.shtml](http://www.boston.com/globe/magazine/8-29/featurestory1.shtml) David Bearman has recently written a thoughtful paper describing models for "intellectual property conservanc[ies]." [http://www.dlib.org/dlib/december00/bearman/12bearman.html](http://www.dlib.org/dlib/december00/bearman/12bearman.html)

How might something like "national parks for intellectual property" work? First, preserves would be repositories for intellectual property rights that had been donated by rights holders. These rights could include copyrights, or in the case of public domain materials, the right to reproduce and
disseminate them. The activities of the preserve would be closely coordinated with existing archives and libraries, which might often still hold physical materials.

Preserves could hold published and unpublished, analog and digital materials of all kinds. Assets could be acquired by purchase or donation. Why would copyright owners (or owners of public domain materials) ever cede their properties to the preserve? First, and perhaps most important, are tax incentives. These would require amendments to the tax code to allow substantial deductions or tax credits for donating valuable copyrights or materials. Second, following the precedent of public land acquisitions, key donors might be compensated with private funding. Third, recognize that donations are prestigious deeds benefiting the national cultural heritage.

There is nothing particularly radical about the concept and practice of a preserve. It's an attempt to work within the system, a gentle expropriation, a creation of incentives for property holders to do the right thing. Ultimately, though, its goals are to rebalance private vs. common property for mass benefit. The preserve aims to make a significant portion of our intellectual and cultural property available to one and all -- both individuals and corporations -- for nothing more than the physical costs of duplication and transmission, and in some cases that the local library might absorb that cost. Its concept supports freedom of inquiry and freedom of expression by implementing access rights in the broadest sense -- to quote, to duplicate, and to appropriate preexisting material.

**PRESERVES -- A CASE STUDY**

At the Internet Archive, we are moving in the direction of building an IP preserve by creating a new paradigm for access to archives, in this case historical film from Prelinger Archives, a private collection of 45,000 ephemeral films physically located in New York and San Francisco. What we've done is to select a body of 1001 key archival films, films we have found to be most in demand, plus unknown films that experience suggests people will want to see and work with. Next, we've undertaken the expensive process of transferring the films to videotape and then digitizing them so that they can be stored and served online (http://www.moviearchive.org).

We offer these digital video files in two resolutions: the fairly high-resolution MPEG-2, suitable for exhibition, rebroadcast, and incorporation into other productions; and the smaller, lower-resolution MPEG-4, suitable for research, reference, and more casual viewing. All files are downloadable only. We have ruled out offering streamable video for technical considerations and out of a belief that users should have the right to view material repeatedly and do with it what they please. Stills grabbed from the films are also available.

In the two months that our server has been up, over 10,000 movie files have been downloaded. Though the sample is still too small to analyze definitively, the preponderance of downloads are from educational institutions and DSL/cable modem users, as might be expected given the large size of the movie files (approx. 250 MB/10 minutes). As more films are converted into MPEG-4 files, which are about 12% of the size of corresponding MPEG-2 files, we expect many more downloads from a more diverse group of users.
Since all of the material is either owned by Prelinger Archives or in the public domain, user restrictions are minimal. First, users are asked not to resell the data files themselves, or use them to go into the stock footage licensing business. Second, they are encouraged to propagate the material, but in all cases we ask that money not change hands. Third, they are asked not to convert the films into closed-source formats.

One key issue -- the effect of this initiative on the donor archive -- remains open. Will the availability of 1001 key films from Prelinger Archives cause significant losses in stock footage licensing revenue, or will the concomitant publicity put the archives on the map and increase business? Ubiquity of an image has never been a deterrent to stock photography sales, just as repeated airplay has never reduced the value of a music copyright. It seems likely that high-end users will demand the best quality imagery available rather than settling for compressed video, but this is not yet clear. What is clear, though, is that Prelinger Archives now has a long-desired means of providing mass access to a subset of its holdings without having to subsidize the costs of such access. In effect, it trades the risk of losing a measure of control over key materials for the benefit of being able to provide cost-free access. And there is no doubt that the number of "access events" will be extremely high.

Already several users have responded to our initiative by offering to donate film or video of their own to the online collection. We hope that institutions as well as individuals will join this effort by agreeing to relinquish total exclusivity to the content they hold in exchange for the visibility and reach that open access can provide them.

By propagating a canonical set of historical films, we also hope that popular authorship will expand to accommodate time-based media. While many people write, make visual art, compose and play music, relatively few people make videos and films. Is the comparative lack of work in moving image media a result of lack of imagination, or tool complexity, or want of opportunity? Until recently, authoring tools were rare and expensive, and at least some of the problem has resided in the lack of access to content. Very few people have ever had significant access to primary moving image historical material, so it has had little chance to seep back into, and influence, the culture. There is much talk of media literacy, but access to the archival material that might make exercises in literacy dig deeper has always been difficult. Here is an initial step toward making it easier.

While it may require the support and expertise of elite circles to organize something like an intellectual property preserve, and though the activities and even the existence of such a preserve might be invisible to most people, a preserve could mount a fundamental challenge to our definitions of public and private property. In so doing, it would be a greater force for change than any possible reform of copyright law. The Internet Moving Images Archive is a small, but functional, step in this direction.
Interlibrary Loan as a Model for Digital Materials

Interlibrary loan (ILL) is the process by which a library requests material from, or supplies material to, another library. ILL expands access to needed material by providing a system that permits libraries to share books, microfilm, and other returnables as well as supply copies of journal articles, book chapters, and other non-returnables. Although some librarians use the phrase “document delivery” to describe the requesting and supply of copies, the more encompassing term of ILL is used to include both returnables and non-returnables. The ILL process has traditionally been mediated. That is, the library, acting on behalf of its user, requests an item from another library, receives the item, lets the patron use the item, and then returns the item to the owning library.

American libraries have shared materials with each other since 1876 when the director of the Worcester, Mass. Free Public Library suggested that libraries lend books to each other for short periods of time. The requesting portion of the ILL process was automated in the late 1970s with the introduction of the OCLC and RLIN ILL messaging systems, which now use the Internet rather than proprietary networks to transmit ILL requests. Libraries began using fax machines in the mid 1980s to expedite the shipment of photocopies. In the early 1990s, the Research Libraries Group introduced the Ariel software to permit libraries to scan copies of journal articles or book chapters and send those scanned images via the Internet to the requesting library. The 2001 revision of Interlibrary Loan Code for the United States no longer limits what can be requested on ILL to scholarly or research materials, as was the case with earlier codes.

In today’s environment, the ILL process is still mediated by library staff. Patrons submit requests to their ILL department, and ILL staff confirms the accuracy of the information, select one or more potential lenders, and order materials on behalf of their users. The library that owns the item sends the item to the requesting library via the U.S. Postal Service, commercial carrier such as UPS, or a state or regional courier. For photocopies, libraries send the copy by mail, fax, or Ariel. Within the past year requesting libraries have begun using Prospero, free software that converts Ariel documents into PDF files and then mounts them on a secure web server. Prospero emails a notice to patrons informing them of the URL where their documents are temporarily stored.

The legal basis of interlibrary loan is grounded in the U.S. copyright law. The borrowing portion of the ILL process is supported by Section 108(g)(2). This section permits libraries to request copies of articles for their patrons as long as they do not use ILL to avoid purchasing the book or journal. Section 108(g)(2) is further supported by the CONTU Guidelines that describe the phrase, “in such aggregate quantities” used in that section. The CONTU Guidelines permit a borrowing library to receive five articles from the most recent five years of a journal title. The CONTU Guidelines do not have the force of law, but are widely followed by libraries.

The lending portion of the ILL process is supported by the first sale doctrine (Section 109) that permits libraries to purchase books and other materials and lend them to their own patrons or to other libraries. The lending portion is also supported by Section 108, which permits libraries to make copies under certain circumstances such as preservation, replacement, or ILL. Libraries may make copies for another library as long as the copy becomes the property of the user and the library
filling the ILL request has no indication that the copy would be used for any purpose other than private study, scholarship, or research.

Historically libraries have been able to collect the majority of materials needed by their local clientele. As a result, the volume of ILL borrowing has been modest. ILL borrowing is less than five percent of a research or college library’s circulation total. Because of this modest volume, most research and college libraries have absorbed ILL costs. Medical and special libraries also view ILL as an essential service, and thus do not pass along ILL fees to their users. However, most public libraries do not budget for ILL fees. As a result, many public library patrons are asked to pay fee when they submit ILL requests, and many of those patrons are also asked to pay the lending library’s loan or copy fee.

As lenders, many libraries (research, college, medical, and special) have imposed charges to lend a book or supply a photocopy; lending fees charged by public libraries are less common. Most ILL fees are transaction-based. One interesting variation to transaction-based charging is the Center for Research Libraries (CRL). CRL was formed by a group of midwestern universities to provide remote storage for lesser used material and has evolved into a collection of specialized materials selected by members. Libraries join CRL and commit to help build and maintain CRL’s collections. One of the benefits of CRL membership is the ability to request any item held by CRL via ILL. The cost of filling those ILL requests is built into the CRL membership fee. Non-members are limited in the number of ILL requests they can submit and they are required to pay a lending fee of approximately $100 per request. The variety and depth of CRL collections is an incentive to join, but even the Center does not house all materials needed by its members.

The interlibrary loan model offers an attractive model for providing access to digital materials. For the past decade, libraries have used digital technologies to share print-based material. Libraries have used and continue to use digital fax machines or Ariel workstations to scan and send articles and other excerpts from print-based materials. Because contract law, not copyright law, governs electronic (born-digital) resources being increasingly acquired by libraries, libraries are not able to rely on Section 108(g)(2) to make and send copies to fill ILL requests. The license must specifically permit libraries to use the licensed digital resource to fill ILL requests, but not all licenses provide the same level of use as currently supported by the U.S. copyright law and CONTU Guidelines.

Copyright issues aside, one of the remaining challenges of using the ILL model as the basis of access to digital resources is determining where the needed item is located. Staff in ILL departments currently spend a significant amount of time searching OCLC and/or RLIN union catalogs to find libraries that own the item. ILL staff also search a variety of other sources, including Web-based library catalogs, an effort that is not sustainable as more material is found only on the Web. Having an universal, comprehensive, and single point of access to all digital resources is an essential component for this model to succeed.
In summary, the ILL model for providing universal access to cultural heritage would encompass the following characteristics:

- The process of requesting and receipt would be mediated by library staff;
- The requesting library would have limits on the quantity of what it can request;
- The digital copies would become the property of the user;
- The process would not serve as a substitute for an item if the level of local use was sufficiently high; and
- The range of materials would be greatly enhanced to include print and digital resources.

## Lending Digital Materials to the Public

Public libraries acquire large collections of books and other works at full price and then lend them out. Ben Franklin and Andrew Carnegie put this in action in the United States to promote democracy, education, and the advancement of the underprivileged. As a society, we have supported these values in our laws and our tax dollars. Six billion dollars was spent in 1997 on public libraries, of which 15% were spent on acquiring the collections. Where printed objects must be lent out physically, the digitized and digital library holdings can be lent out digitally. Using this model, libraries can continue to serve the public good in the future without a major change to the institutional structure as we digitize the collections.

This model is different from the interlibrary loan model because it assumes that the local library owns or licenses the needed material.

Several technologies are now available to allow lending, one is netLibrary which is a technology and a service, another is eBook by Adobe Inc (see appendix 1) which is just a technology, and a public domain technology called “Lending Library Format” that is starting to be used for a few collections. These copy protection technologies are never ironclad since if a file is viewable it is recordable, possibly with some degradation in quality. On the other hand, library patrons can copy things from physical libraries as well, and have been doing so both legally and perhaps illegally since libraries began.

The San Francisco Public Library system has started lending out electronic books to anyone that holds a SF library card by using netLibrary.com. San Francisco Public determines how many copies of a particular title should be made available and the loan period for each. Using a library card, a user can search and download books to their hard drive. If the library had selected only one copy, another SFPL user would be told that the title is in use. After 3 days, if that is the “due date” set by SFPL the user is not able to open the file and is prompted to check the book out again.
After a netLibrary book has expired the user who tries to launch the book gets this notice:

```
'The Copyright Book' has expired. Would you like to attempt to check out another copy from netLibrary?

[ ] Yes  [ ] No  [ ] Cancel
```

While there are not very many books in this format currently, most patrons will understand the procedure because it is closely modeled on checking a book out of a library. Rather than returning the physical book to the library, the electronic equivalent is for netLibrary to disable the copy on the hard drive.

Other software publishers are making technologies that can be used in such systems.
A public domain system for lending digital files has been written called the Lending Library Format, which is being tested with limited collections. With this system, a library can lend a digital file in much the same way as netLibrary, but without having to have a central company prepare the material. With this capability, libraries could digitize rare and fragile holdings or record materials only available in digital form and then lend them out. Because of the reduced cost of digitization or recording digital files, comprehensive collections could be made available in the near future.
Where lending physical copies guarantees that only one person has a book at a time, achieving the same restriction with digital files requires extra work. The 1976 US copyright law made provisions on how this can be accomplished for one electronic collection, television news programs, and put in provisions of what lending such materials means (see sections 108(f)3 in http://www.loc.gov/copyright/title17/92chap1.html#108).

Therefore the analogue to lending physical materials is to lend digital materials. Libraries would then make their digitized and digital materials available to their patrons. Audio, video, scanned books and periodicals, scanned photographs, and ephemera could be made available with limitations on who can have access to these collections and for how long. Even beyond what is possible with physical objects, different restrictions can be imposed such as username/password, library card number, IP address, etc.

Digitization prices have dropped, so that making unique collections available from libraries is becoming more affordable. Scanning a page is now under $0.10 even in fairly small quantities of materials (e.g. http://www.archive.org/arpanet and http://www.roottech.com/), scanning an hour of videotape is about $10 (e.g. http://www.archive.org/movie), recording an hour of a radio broadcast costs about $1.

The cost of serving digital materials is also dropping if libraries are careful in what they purchase. A terabyte server that cost under $7000 (e.g. www.alexa.com and http://www.aslab.com/) will store the text of one million books or about one thousand hours of video.

Therefore libraries can bring parts of their own collections online cost effectively to serve their patrons.

By borrowing a work in Lending Library Format, the library patrons would use a system similar to netLibrary, but instead of using a custom reader, it would use the normal reader for that file format. The reason this would work is that the Lending Library Format is an “envelope” around a file that indicates the start and stop times it is legitimate to view the file. If it is within that period then a simple client program decodes the file and passes it to the appropriate viewer, if it is not within that period it prompts the user to check the work out from the library again.
Conclusion: Roles Rights and Responsibilities

Since the roles and responsibilities of libraries for providing public access have not changed because of digital materials it is reasonable to assume that the rights to perform these societal tasks have not significantly changed either. Leveraging the existing system of library and archives with its existing approaches can be easily extended to include digital materials. A combination of the three structures discussed here—donations, interlibrary loan, and lending libraries—have served well in the past and can in the future. What will be required are leaders in building examples of the technologies and collections. These leaders in libraries, archives, and public policy groups can have a far-reaching impact on the future of education and scholarship by offering universal access to our digital cultural heritage.
Related Work and Related Organizations

Prelinger Archive http://www.prelinger.com

How much information is there?
  Lesk: http://www.lesk.com/mlesk/ksg97/ksg.html
  Lyman and Varian: http://www.sims.berkeley.edu/how-much-info/
  Library of Alexandria, interesting book: Canfora, Vanished Library

Web Archives:
  Internet Archive http://www.archive.org/
  Google’s web cache is a form of short-term archive

Text Archives on the net:
  Project Gutenberg http://www.gutenberg.org/
  University of Virginia http://www.lva.lib.va.us/dlp/
  CMU Universal Library http://www.ul.cs.cmu.edu/
  Oxford Text Archive http://ota.ahds.ac.uk/

Notable Large Archives of General Interest:
  Archive of Contemporary Music www.arcmusic.org
  Center for Research Libraries wwwcrl.uchicago.edu
  Internet archive www.archive.org
  Music archive www.musicarchive.org
  National archives and Records Administration www.nara.gov
  Television News archive tvnews.vanderbilt.edu
  Television Archive http://www.televisionarchive.org/

Library Associations and Organizations:
  American Library Association www.ala.org
  Association of Research Libraries www.arl.org
  Coalition for Networked Information www.cni.org
  Council on Library and Information Resources www.clir.org
  Digital Library Federation www.clir.org/diglib/dlfhomepage.htm
  Online Computer Library Center www.oclc.org
  Research Libraries Group www.rlg.org
  Special Library Association www.sla.org
  Medical Library Association
  Music Library Association
  American Association of Law Librarians
Law and Policy Centers with Library Interests
Electronic Frontier Foundation http://www.eff.org/
Samuelson Law, Technology and Public Policy Clinic at UCAL Berkeley
   http://www.law.berkeley.edu/news/releases/20000424Samuelson.shtml

Technologies and Systems:
   NetLibrary  http://www.netlibrary.com/
   Adobe eBook  www.glassbook.com

Economic model for shared resources:
   “National Public Radio” model: John Willinsky from U of British Columbia
http://www.educ.ubc.ca/faculty/ctg/pkp/kem/kem.pdf

Appendix 1

from (http://www.pigdogs.org/art/adobe.html)

Adobe’s eBook reader tries to technically confine the user to obey the restrictions proposed by the publisher. It uses technology to try to stop the cut-and-paste feature of operating systems, stop running debuggers, and warningreaders from disobeying a publishers interest. In this screen shot, for instance, it declares that reading Alice in Wonderland aloud is not allowed.
First published in 1865
by Macmillan & Co., London
Released 1866 by D. Appleton & Co., New York

For information
about VolumeOne and unit-run printing, contact:
Peter Zelchenko (pcete@chinet.com)
1757 W. Augusta Blvd.
Chicago, IL 60622-3200 USA
(312) 733-2473

The text of this book was originally entered as an online etext for Project Gutenberg™ and was subsequently prepared for print publishing by the VolumeOne staff. VolumeOne is grateful to Project Gutenberg for its contribution to this work, VolumeOne holds harmless and indemnifies Project Gutenberg of any liability arising from the use of their text in this printed embodiment.

Text from Project Gutenberg
For more information on Project Gutenberg, contact:
Project Gutenberg
Michael S. Hart (hart@vmd.cso.uiuc.edu)
P.O. Box 2782
Champaign, IL 61820

Permissions on : Alice’s Adventures in Wonderland

**Copy**
No text selections can be copied from this book to the clipboard.

**Print**
No printing is permitted on this book.

**Lend**
This book cannot be lent or given to someone else.

**Give**
This book cannot be given to someone else.

**Read Aloud**
This book cannot be read aloud